Invisible genomes: the genomics revolution and patenting practice

Invisible Genomes:

The Genomics Revolution and Patenting Practice

Published as: Bostanci, A., & Calvert, J. (2008). Invisible genomes: the genomics revolution and patenting practice. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 39(1), 109-19, doi: 10.1016/j.shpsc.2007.12.010

Adam Bostanci*
ESRC Centre for Genomics in Society
University of Exeter
a.w.s.bostanci@ex.ac.uk

Jane Calvert
ESRC Centre for Genomics in Society
University of Exeter

Abstract

In the mid-1990s, the company Human Genome Sciences submitted three potentially revolutionary patent applications to the U.S. Patent and Trademark Office, each of which claimed the entire genome sequence of a microorganism. The patent examiners, however, objected to these applications, and after negotiation they were eventually re-written to resemble more traditional gene patents. In this paper, which is based on a study of the patent examination files, we examine the reasons why these patent applications were unsuccessful in their original form. We show that with respect to utility and novelty, the patent attorney's case built on an understanding of the genome as a computer-related invention. The patent examiners did not object to the patenting of complete genome sequences as computer-related inventions on moral grounds or in terms of the distinction between a discovery and an invention. Instead, their objections were based on classification, rules and procedure. Rather than patent examiners having a notion of a genome that should not be patented, the notion of a 'genome', and the ways in which it may be different from a 'gene', played no role in these debates. Our findings have far-reaching consequences for patenting in the biosciences.

Keywords: Genome; Genomics; Bioinformatics; Intellectual Property Rights; Patent; Classification

Introduction

The esoteric realm of patent practice might, at first glance, seem a strange place to look for reflections about the nature of biological entities like genomes. During the process of patent examination, however, decisions have to be made about whether such entities are eligible for patent protection. Views have to be formed about the nature of these entities and whether they qualify as inventions. We became interested how the entity called 'the genome' was seen in patent practice. More specifically, we were interested in whether patent examiners perceived genomes differently from genes, which have been routinely patented since the 1980s.[1]

This question was raised for us by the discovery of three patent applications filed by the company Human Genome Sciences in the mid-1990s. These applications were unusual in that they not only sought to claim as an invention the complete genome sequences of three microorganisms (*Haemophilus influenzae, Mycoplasma genitalium*, and *Methanococcus jannaschii*), but they claimed these genome sequences in computer-embodied form. By the time patents were granted for these applications, however, the invention claims had been modified to cover only a small number of genes. We wanted to find out why the scope of these patent applications was restricted in this way, and whether arguments about the patentability of genomes played a role in this process. At the time, biology was undergoing a transformation due to the rise of computing, the internet, and bioinformatics, so these patent applications raise the question of whether a genome can be understood as a computer-related invention.

It is easy to speculate about why these patent applications might have been rejected. Intuitively, genomes are conceived as intimately connected with, or even constituting, the nature of an organism and are therefore presumed to be unpatentable (see Blute, 2005). This view has fuelled broader political and moral debates about the desirability of human gene patenting. In addition, Eisenberg (2000), who has offered the only scholarly discussion of these patent applications, has argued that genomes ought not to be patented in this way because a genome sequence is 'scientific information' that constitutes a basis for future scientific discovery.[2] She also makes the point that while DNA sequence information may well have commercial value, allowing such claims would amount to a back-door extension of a patent system that was originally designed for 'bricks and mortar' inventions to intangible inventions, an extension that has not been subjected to proper policy or legal analysis.

These intuitions notwithstanding, a previous survey has found that a number of patents for whole genomes have nevertheless been issued, albeit primarily as patents for biochemical entities with a view to the pathological specificity of particular microoganisms (O'Malley *et al.*, 2005).[3] While Eisenberg has provided an argument for why genomes ought not to be patented in computer-readable form, we wanted to explore why they were in practice not deemed patentable in this way by the U.S. Patent and Trademark Office (USPTO). In brief, we found that computer-embodiment was indeed part of the patent attorney's case for patenting the whole genome, whereas patent examiners failed to see the genome as an invention because of the rules that govern patent examination and because of the patent classification system.

We begin this paper by providing an overview of the patent system and of the patent examination process as they emerge from our case study. We then review the scientific history of the three genomes and the patent applications, making reference in particular to Cook-Deegan's (1994) argument about the inseparability of computing from genomics. In our analytical section, we show that computer-related arguments figured in the case the patent attorney made for the patentability of the genome. A utility for the genome as a computer-related invention was articulated, although, in contrast to gene patents, this utility did not coincide with the presumed biological function of the genome. Also with respect to novelty, we find that computer-embodiment was part of building a case for the patentability of the genome. We then consider reasons why the genome as presented in the patent applications remained 'invisible' to examiners. Here, we discuss the all-important classification scheme employed by the patent office. The tripartite nature of the U.S. patent classification system—mechanical, chemical, and electrical—and the requirement of 'restriction' means that the applicants were forced to restrict their invention claims to selected genes only. This not only rid the applications of their computer-related claims, it also meant that they had to give up on patenting a complete genome. We end with some thoughts about role of patent classification and its implications for the future patenting of biological entities as biological research moves towards systems biology and nanotechnology.

Methodology

The scholarly literature on intellectual property usually analyses laws, cases and judgments. Others have studied how scientists and patent professionals produce texts (Myers, 1995; Packer and Webster, 1995; 1996). In contrast, this paper is based on an analysis of patent

examination files. We used the online Patent Application Information Retrieval system of the USPTO to identify the patent applications that gave rise to the most correspondence between patent attorneys and patent examiners from the patent families associated with the three genomes.[4] The patent examination files contain all correspondence leading up to a patent being issued, and they are typically 1000-2000 pages long, although much of the file is taken up with a print-out of the complete genome sequence.[5] By studying the file histories, we see how decisions are made during patent examination. In addition, this analysis draws on the scientific articles describing the complete genome sequences and on interviews with scientists involved in the production of these genome sequences as well as the patent attorney who prepared the genome patent applications.[6]

Patenting Genomes

A patent is a contract between government and inventor, preventing others from making, using or selling the invention for a period of 20 years. In turn, the inventor makes the invention available for the benefit of the public. In the US, the basic conditions for patentability are that the invention is novel, non-obvious (to someone 'skilled in the art') and useful. Generally 'anything under the sun made by man' may be patentable (see Kevles, 2002, p.25), although 'laws of nature', 'abstract ideas', and 'products of nature' are explicitly barred from patent protection. Since the three patent applications we analyse here were originally filed and examined at the USPTO, the primary research presented in this paper mainly reflects U.S. patent practice. We also studied the examination of genome patent applications at the European Patent Office for comparative purposes.

The patentability of DNA and inventions in the field of molecular biology has been firmly established since the *Chakrabarty* decision of the Supreme Court in 1980, where a patent was granted on a modified bacterium (Kevles, 2002). In this case, it was held that biotechnological inventions were eligible for patent protection due to their analogy with chemical inventions. Following the *Chakrabarty* case, many different kinds of patents in molecular biology were issued. Precedent was set that 'isolated and purified' nucleic acid molecules, such as a human gene excised from a chromosome, are patentable like other isolated and purified chemical compounds. However, 'gene patenting' has been mired in controversy. For example, patent applications for partial human gene sequences have been contested with respect to their utility within the scientific community (Cook-Deegan, 1994; Marshall, 1994). More recently, human gene patents have given rise to debates that are often framed in moral terms and derive their

force from the idea that the genome constitutes the nature of the organism (Sulston and Ferry, 2002).

The patent applications for the complete genome sequences of *H. influenzae*, *M. genitalium*, and *M. jannaschii* were filed in the mid-1990s, a time when DNA sequencing capacity was growing rapidly, due to the availability of automatic DNA sequencing machines. The ability to sequence whole genomes led to an exponential increase in the number of DNA-related patent applications filed at the USPTO (Marshall reported a ninety year backlog of DNA-related applications in 1996). Partial human gene sequences in particular were often covered by blanket patent applications containing a large number of DNA sequences. These developments were addressed by new bureaucratic measures. Most importantly for this paper, a rule was added to the Manual of Patent Examining Procedure (MPEP) that no more than ten DNA sequences could be claimed in a single patent application – a simple measure to manage the workload of the USPTO (USPTO, 2001).

As well as being influenced by gene patenting, the genome patents discussed in this paper were influenced by another contemporary development, the patenting of computer-related inventions (interview Millman). Such inventions were not considered patentable until the mid-1990s when, following a series of judgments, the USPTO developed ‘Examination Guidelines for Computer-Related Inventions’, which set out that software can be patented in conjunction with a specific machine as a 'programmed computer', but excluded computer media storing 'non-functional descriptive material' such as music (USPTO, 1996). These guidelines instruct patent examiners on how to assess patent applications for 'computer-related' inventions.

Patent examination involves a patent examiner, acting for the government, and a patent attorney, acting for the inventors. In the first step, the examiner reads the two parts of the patent application: the 'specification' and the 'claims'. The former provides a general discussion of the invention, while the latter summarises what the applicants see as their specific inventive achievements (Figure 1). Once a patent is granted, the claims are normally taken to define the scope of protection given by the patent. Patent professionals often treat the specification as largely inconsequential, but in the initial stages of patent examination, the specification is arguably more important than the claims, because it gives the examiner an impression of what the applicants believe to be their invention. Much of our argument below

draws on the patent specifications. Subsequently, the examiner identifies the relevant patent classes in U.S. Patent Classification system and conducts a 'search' of 'prior art' to assess whether the invention fulfils the criteria for patentability (USPTO, 2005). As with any other contract, the next step is normally a negotiation between the examiner and the attorney to establish the precise wording of the invention claims.

A lesser-known aspect of patenting is 'restriction'. If, after the initial reading of the patent application, the examiner comes to the view that the application in fact encompasses several different inventions, he can demand 'restriction' to a single invention. Restriction can be required if more than one 'independent and distinct' invention is claimed per application and if search and examination of these inventions would impose a 'serious burden' on the examiner (USPTO, 2001). The purpose of this rule is to prevent inventors from protecting more than one invention in a single patent, and to reject applications that overwhelm the examiner with complex claims that require searches in multiple fields of prior art. Clearly, it would be impractical and impossible to search all previous inventions described in some seven million patents for each new patent application. Since misclassification threatens to confound future searches for prior art, consistent patent classification is in the interest of the USPTO examiners. As we argue in our conclusion, the importance of classification in the work of a patent office can hardly be overestimated.

Patent examination emerges from our case study as a rather innocuous process, governed by Manual of Patent Examining Procedure, the U.S. Patent Classification, as well as specific guidelines, for example for 'computer-related' inventions. However, given the technical nature of inventions and the requirement that inventions be novel, one would expect that the distinction between one and several inventions is not always clear-cut. While patent law has procedures for dealing with difficult cases, patent practitioners have pointed out that restriction is difficult to understand, operates imperfectly and would benefit from simplification (Henry, 2004). All this matters because for the applications for the genomes of *H. influenzae*, *M. genitalium*, and *M. jannaschii*, one of the most important arguments between the patent examiners and the patent attorneys was over whether restriction was required or not; that is, whether the patent described one or several inventions.

This leads us to a final aspect of patenting practice: unlike legal arguments in court, the patent examination process does not encourage lengthy arguments. MPEP constrains the outcome of

disagreements. If the applicants disagree with the examiner they may set out a counter-argument but at the same time they are obliged to choose one of the inventions identified by the examiner for further examination. For this reason, the disputes over the patentability of the three genomes we are analysing here gave rise to arguments that may be described as succinct or perhaps even cryptic. Extensive textual analysis was required to bring to the fore what the key argumentative exchanges consisted in and why the patent attorneys failed to persuade the examiner to view the genome as a patentable invention. Although we expected to find arguments about the nature of the genome in the patent application files, the exchanges between applicants and patent examiner hardly considered the nature of the genome at all. Instead, the genome, as articulated in the patent application, was all but invisible to the patent examiner.

The Scientific History of the Genomes

The three genomes were sequenced at The Institute for Genomic Research (TIGR), a not-for-profit organisation headed by Craig Venter, who later became CEO of the company Celera Genomics. At the time, TIGR was bound by a contractual relationship with Human Genome Sciences, which gave the company a period of six months to secure intellectual property protection for all TIGR discoveries. This arrangement led to a dispute over the publication of the first of the three genomes, that of *H. influenzae*. While TIGR scientists wanted to publish a scientific paper describing the genome, the company wanted to ensure that no potentially valuable intellectual property was forfeited, and for each of the three genomes a patent was filed before the publication of the sequence data to ensure priority (Shreeve, 2004). The patent application for the *H. influenzae* genome then served as a template for those subsequently filed for *M. genitalium* and *M. jannaschii*, and broadly similar argumentative exchanges between the applicants and the examiner are found in the three files. For this reason, we discuss all three file histories together.

Once published (Fleischmann *et al.*, 1995), the genome sequence of *H. influenzae* was hailed as a milestone in the history of biology for being the first complete genome sequence of a self-replicating organism (Nowak, 1995; Wade, 1995). Whereas previously it had been possible to determine the DNA sequences of particular genes or the genome sequence of simple viruses, automated DNA sequencing machines along with the data-processing powers of computers now made it possible to determine the DNA sequence of an entire microbial genome. The analysis involved further computationally-aided comparisons of the genome

sequence with the DNA sequences collected in databases such as GenBank, and identification of genes and their functions on the basis of DNA sequence similarity.

The choice of organisms illustrates the broader biological ambitions connected with this research. While *H. influenzae* was simply the experimental organism of a laboratory associated with TIGR, the other two organisms were chosen more strategically. It was expected that the *H. influenzae* genome would contain complete sets of genes known to be 'essential for life'. However, when the genome was analysed by identifying sequence similarities within the genome sequence with database matches to known genes, three enzymes that had been assumed to be universal were missing. For this reason, TIGR scientists decided to next sequence the smallest known bacterial genome, that of *Mycoplasma genitalium,* which at the time was thought to possess the smallest genome for a self-replicating organism and, thus the reasoning, should possess all essential genes (Fraser *et al.*, 1995). The third organism that was sequenced, *Methanococcus jannaschii*[7] is an Archaeon (a type of organism structurally similar to bacteria but genetically different), and it was chosen so that all three domains of life (Archaea, Bacteria, Eukaryota) could be compared. Moreover, *M. jannaschii's* ability to survive without oxygen and to synthesize organic materials from inorganic chemicals made it interesting to biotechnology, astrobiology and biology in general. Scientists hoped that the genome sequence would help understand these biological capacities (Bult *et al.*, 1996). These research projects were motivated by the hope that knowledge of complete genome sequences would provide a more global view of the metabolism of the organisms and perhaps even an understanding of the genomic basis of self-replicating life.

In his influential account of the early history of the Human Genome Project, Cook-Deegan (1994) argues that, at the time, "computers, and mathematical algorithms [were] as important as DNA sequencing, cloning, and other more obviously biological techniques," (p.288) and that advances in computing were an "essential element in the pursuit of genetic knowledge" (p.291). Discoveries could be made "simply by comparing [sequence] data in different databases" (p.293). He also shows that it was still unclear whether software development to support this research would be controlled by scientists or private vendors. Arguably, knowledge of the three genomes discussed above, achieved through elucidation of the complete sequence and subsequent analysis by means of computer-aided database searches, depended on advances in computing as highlighted by Cook-Deegan. As we will see, the patent attorney representing Human Genome Sciences made a similar argument.

Patent Applications

As indicated above, the reason why these patent applications were so unusual was the emphasis on the genome as a computer-related invention. For example, the first invention claim in the *H influenzae* and *M. genitalium* genome patent applications reads: 'Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID NO:1'. SEQ ID NO:1 represented the entire sequence of the genome, which was defined as the 'life sustaining instructions and information' of the organism (USPTO Appl. Nos. 08/476,102; 08/545,528). Influenced by the emerging patentability of computer-related inventions, the patent applications even include a diagram of a computer-system with a storage medium carrying the genome sequence (Figure 2), and for the *H. influenzae* genome, the field of invention was changed from 'molecular biology' to 'bioinformatics' after the application was filed (Examination File, Amendment A, 12 September 1995). The patent specifications explained that the genome sequence in computer-readable form would serve for homology searches (i.e. comparisons with other DNA sequence databases), with the aim of identifying commercially or biologically important fragments of the genome. The applications also mentioned that knowledge of the genome sequence would further the 'understanding of chromosome structure and function', 'the structure, position, and spacing of regulatory elements' and the ability to do 'comparative genomic and 'molecular phylogeny', although these utilities were not expounded in depth.

The applicants not only attempted to claim the genome sequence in computer-readable media, they also attempted to claim isolated fragments of the genome and the proteins encoded by the genome. Arguably, the identification of all these valuable fragments is testament to the utility of the genome in computer-embodied from. Perhaps this can also be understood in the tradition of gene patenting, where a single patent often protects a range of different biochemical entities including the nucleotide sequence, the protein encoded by the sequence, and the antibodies that selectively bind to the protein. If a 'gene patent' can achieve all of this, then it might seem plausible that a 'genome patent' might encompass all of these applications for all the genes of an organism.

These claims gave rise to an argument between patent examiner and attorney. The examiner felt that the applications covered several different groups of inventions and, for *M. genitalium,* he demanded restriction to one of the following groups: I. computer readable medium, II.

computer system, III. methods of using a plurality of nucleic acid sequences, IV. individual open reading frames (i.e. sections of the genome that code for example for proteins, referred to below as ‘ORFs’), V. non-coding nucleic acid sequences, VI. peptides, and VII. antibodies. He argued that these seven different types of invention are unrelated because they have 'different modes of operation', 'different chemical structures, physical properties and utilities'. He argued that because they fell into different patent classes this meant that they had 'divergent subject matter' (Restriction, 29 January 1999). In this rather circular manner, he relied on the pre-existing classification scheme to decide how to separate out the different types of invention.

The patent attorney (acting on behalf of the inventors) objected that restriction to one of these groups was not required because search and examination would not be unduly onerous. The records in the file indicate that, in making this argument, he may have drawn on his experience with *H. influenzae*, where electronic searches with the term 'genom?' in the database of the USPTO during the preparation of the patent application had revealed no countervailing prior art (Amendment A). Perhaps encouraged by this experience, he argued that searches of the different groups would not be a burden, explaining that in most publications 'where a published polynucleotide is shown, the authors also include the computer readable medium, a computer system, polypeptides, antibodies, and methods using a plurality of nucleic acid sequences'. Like Cook-Deegan, he suggested that in genomics advances in computing and in molecular analysis went hand in hand to such an extent that 'the evolution of one is tied to the existence of the other' (Amendment C, 5 March 1999).

But the examiner remained adamant, maintaining that 'computer hardware is an *entirely different technology* from molecular biology' (emphasis added), and that searches for 'nucleic acid molecules would never uncover art related to computer systems'. For him the fact that 'all the [genes] are presented as part of one SEQ ID', that all the DNA sequences being claimed were combined in one long sequence, was merely a matter of presentation, they were still 'unrelated' nucleic acid sequences (Rejection, 30 December 1999). He saw no necessary relationship between the different genes inherited in one chromosome. Consequently, the examiner was overwhelmed by the number of searches required for all the possible fragments mentioned in the application. He saw the genome patent application as if it were an application for a large number of gene sequences. As a result of these objections, the applicants restricted their claims to category IV above – ten different genes. This shows that

the genomes were disqualified from patent protection because of the restriction requirement rather than on the grounds of any principled objection to the patenting of genome sequence as biological information. Restriction was the most significant move in the examination of these patent applications, because this forced the applicants to drop their claims for the whole genome, and subsequently the patent applications became more conventional. These more conventional patents were eventually issued approximately eight years after the original applications were filed.[8]

Discussion

1. The Genome as a Computer-related Invention: Function and Utility

In the original patent applications, the patent attorney described the genome sequence as the 'life sustaining instructions and information' of the organism. This view of the genome ties into more general and popular analogies between information and DNA (Kay, 2000). One patent practitioner, unconnected to our case study, even reversed the analogy to argue for the patentability of software. He suggested a 'programmed floppy disk or other storage device performs much of the same role for the machine in question, e.g., a computer system, as a DNA sequence does for a cell. Both encode information used in controlling and directing the operation of a cell or of a computer system' (Toedt, 1995).

However, as philosophers of biology have pointed out, it is problematic to think of the genome as biological information, or of DNA as the 'master molecule' that determines all organismal functions. Thinking in this way can result in traits like pathogenicity and thermostability being attributed to DNA, when these traits are now thought to emerge at higher levels of biological organisation. Although Godfrey-Smith allows that the idea of 'genetic coding' can justifiably be used to describe the role of DNA in specifying protein structure (Godfrey-Smith, 2000; Neumann-Held, 2006), Griffiths (2001) has argued that it is a mistake to move from the notion of the genetic code to general views of organismal processes as information encoded in the genome sequence. The general problem is that, both in terms of structure and function, the genome is not easily separated form the cellular and organismal context, and it is not clear to what extent our understanding of the organism is aided by imagining its genome as disembodied information (Sarkar, 1996).

The informational view of the genome as a discovery system articulated in the patent applications is more interesting. By transferring the DNA sequence to computer-readable

medium the sequence was transformed into a discovery system that could be set in relation to other DNA sequence databases to identify functional fragments. The computer-embodiment made the genome useful (Shreeve, 2004; interview Millman). Crucially, the utility of the genome in this case is substantially different from the utility we find in gene patents, which is usually articulated by referring to the biological function of the gene. The utility of such computer-based homology searching does not coincide with the presumed biological function of a genome: the 'programming' of a cell. However, the envisaged utility of the genome sequences of *H. influenzae*, *M. genitalium* and *M. jannaschii* as computer-related inventions for homology searching was not formally examined at the USPTO because restriction, the examiner's stipulation that the application concerned different kinds of inventions, took place prior to examination based on the criteria for patentability. Examination of the files shows that these utility claims were not pursued beyond the initial application. The fact that these genome applications were unsuccessful therefore does not imply that a genome sequence *in silico* lacks utility, as Eisenberg assumed.

A more radical interpretation of these patents, albeit one that the examiners never considered and the attorney did not develop explicitly, is that the genome sequence *is* a computer-related invention in the sense suggested by Cook-Deegan. If we set aside the *a priori* view that the micro-organismal chromosome is biological information, then the transformation of DNA, falling into the realm of chemistry, into a computer-readable sequence, falling into the realm of electronics, can indeed be seen as an inventive technological achievement that depended on advances in computing. Global views of genome sequences depend on computer capacity to process large amounts of DNA sequence data. Without the aid of computers, the genome sequence, several thousand pages in length when printed, was not interpretable. The assignment of a function, or at least of a putative function, to each of the genes in the genome proceeded by means of homology searches to identify matching gene entries in existing DNA databases. TIGR scientists similarly recall that it was only possible to publish a description of a genome sequence in a scientific publication once this analysis had been carried out, and a story could be told about the genome (interview White). Without this analysis, the genome sequence data was unintelligible. Hence, an argument can be made that the production of a genome from chromosomal DNA is computer-dependent, as the patent attorney also seemed to suggest.

2. Genomes: wholes and parts

For all three genome patents, parts of the sequence had been made publicly available before the filing of the patent. On first glance, prior knowledge of these parts would threaten to undermine the novelty of the claims for the complete sequences. In this section we explore the relationships between genomes (wholes) and genes (parts), and we show that computer-embodiment can again strengthen the case for the patentability of the genome.

In the case of *H. influenzae*, a total of 59 ORFs were known before the genome project. The applicants 'explicitly disclaimed' these ORFS in their application as far as their molecular claims were concerned and felt no need to reflect on whether prior knowledge of these ORFs would affect their claims for a computer-embodied genome sequence (Amendment A). Since they envisaged using the whole genome as a discovery device, we may assume, employing interpretative charity, that they felt that their invention was not anticipated by a small number of known genes. This position becomes even more credible if we think of the computer-embodiment of the genome as providing insights into characteristics of the genome, such as the relationships between genes. If the emphasis is on properties of the genome, then knowledge of a small number of particular genes would not necessarily imply lack of novelty, although the applicants did not develop a formal argument along these lines.

After restriction, the applicants reformulated their invention claims. They now focused their claims on an 'isolated polynucleotide' that comprises the nucleic acid sequence of particular ORFs. The examiner objected that this could also be taken to also refer to 'isolated chromosomal DNA' that had been described in the scientific literature years earlier. In other words, the examiner pointed out that the way the applicants had phrased a claim for a gene could also apply to an entire chromosome. This was inadmissible for it would rule out the novelty of the invention because isolated chromosomal DNA had been described years earlier. The applicants responded to this objection by adding the word 'fragment' to their claim language. Rather than referring to an ‘isolated polynucleotide’, the amended claims were changed to refer to an ‘isolated polynucleotide *fragment*’ (Amendment D, 12 November 1999). This modification made the claims narrower and more specific so they could no longer be taken to refer to the isolated chromosomes of the three microorganisms.[9]

The word 'isolated' is important here because in patent law a gene only becomes patentable once it has been 'excised from a natural chromosome' and isolated from its natural context.

The necessity, imposed by patent law, to isolate fragments in order to patent them raises problems for genome patenting. By definition, it is not possible to excise a genome from a chromosome, since in molecular terms the genome comprises all of the DNA in an organism's complement of chromosomes. This implies that a whole genome sequence cannot be claimed in the way in which one would normally claim an isolated molecule as a fragment.[10] Indeed, the patent attorney recalled that the computer-embodiment set the genome apart from previous preparations of chromosomal DNA (interview Millman). As a computer-related invention, the genome can immediately be seen as distinct from 'isolated chromosomal DNA', because a molecular entity is transformed into a discovery system embodied in a computer. Hence, the computer-embodiment was also chosen to avoid anticipation of the genome as an invention by prior preparations of purified chromosomal DNA.[11]

3. US Patent Classification and Restriction

Historically, US Patent Classification is organised as a tripartite categorisation of inventions as mechanical, chemical, or electrical artefacts. Perhaps this helps us understand why the patent examiner felt that the applications encompassed different kinds of technology ('Searches for nucleic acid molecules would never uncover art related to computer systems'), even when the patent attorney suggested that in genomics advances in computing and in molecular analysis go hand in hand. During examination, each patent application is in the first instance associated with one particular patent class, and we have already pointed out that consistent classification is integral to the work of examiners, who have to search prior art on a daily basis. While this is defensible as good patenting practice, the picture that emerges from this characterisation of patenting has far-reaching consequences.

The patent applications for complete genomes failed because the patent examiner felt they encompassed different kinds of technology. Arguably, his reasoning – that there are different groups of inventions because they fall in different classes of the US Patent Classification system – leads to a logical regress. One might ask by virtue of what the different classes were established. This question was not raised during patent examination. Moreover, in contrast to the *Chakrabarty* case, there was no impetus to create a new patent class to accommodate the subject-matter described in these patent applications, perhaps because these applications were not seen to mark the introduction of an important new technology. This implies that all future inventions will be seen relative to the grooves of the US Patent Classification. Perhaps the inventions described in the patent applications for *H. influenzae*, *M. genitalium* and *M.*

jannaschii were 'invisible' simply because they did not fit into the established patent classes.[12]

More generally, Mol and de Laet have shown in a discussion of 'the Zimbabwe bush pump' that the definition of a technology can occur at many levels (De Laet and Mol, 2000). What we unreflectively perceive as a simple bush pump can be defined by its parts, by the specific installation of these parts, the hydraulic principles that operate within 'the parts', or even the broader social configuration around the pump. It is the presumption of a unifying character that makes a physical assemblage into a 'bush pump'. With respect to the genome, this approach raises a number of difficult questions. The examiner, acting for the USPTO, perceived the application as merely listing a large number of genes, whereas the patent attorney acting on behalf of Human Genome Sciences provided some tentative arguments for seeing the whole genome as a computer-related invention. Ultimately, in patenting practice it would appear that whether an innovation appears as one or many inventions depends how it is related to bodies of literature, precedents, and perhaps even different legal regimes. In the case of the genome as a computer-related invention, such an entity was not articulated clearly enough. The genome as a computer-related invention was invisible to the patent system, and the patent applications were unsuccessful as originally filed.

Conclusion

Eisenberg cautioned against genome sequence patents on the grounds that patenting genomic information would amount to an unconsidered extension of patent law. Others have mounted a broad range of social, moral, and political arguments against the desirability of patenting biological entities like genes. However, patent examination in the USA has no mechanism for establishing the socio-political desirability of a patent, nor a formal system for deciding when new patent classes should be created for new kinds of inventions.[13] We have shown that patent examiners, relying on the Manual of Patent Examining Procedure and the U.S. Patent Classification system, formed a view on the patentability of the genome only indirectly in following rules and procedure. In fact, one might say that in not recognising the 'genome' as an invention, the examination process at the USPTO “worked” remarkably well, in relation to its own interests to operate on the basis of a conservative classification system for inventive artefacts, which ensures the smooth running of the operation.

In this paper, we have suggested that claims for the genome as a computer-related invention were integral to the attempt to build a case for the patentability of the genome and not

altogether implausible. Structurally and functionally, knowledge of the genome depended on advances in computing. Moreover, as a computer-related invention, the utility of the genome differed from its presumed biological function, and computer-embodiment also circumvented the problem of anticipation of the invention by isolated chromosomal DNA. We might add that the view of the genome as a computer-related invention distinct from chromosomal DNA would remove many of the moral and political concerns associated with patenting inventions derived from biological materials. Incidentally, this implies that we see patents as contracts that protect inventions, rather than bestowing ownership of biological entities, a view that should, we expect, coincide with patent practitioners' understanding of patents.

However, owing to the strictures of the U.S. patent classification system, computer-embodiment was ruled out of court and with it fell the case for the patentability of the genome. Patent examiners were not concerned with the distinction between a natural entity and an invention, or with the broader moral and political debates about the propriety of patenting biological entities like genes. Perhaps one might even say that the rules and procedures were applied to such an extent that the examiner lost sight of the genome as a possible computer-related invention. Examiners insisted on the categories established by precedent and refused to grant a patent on a genome; but this is because they saw a 'bag of genes' where the applicants tentatively articulated a 'genome'. Our examination of the files containing the to and fro between patent attorney and examiner shows that, paradoxically, a genome as seen through the eyes of the patent office is too many inventions, that is genes, rather than not inventive enough.

These finding have far-reaching consequences for biology-related inventions. The increasing practical utility of biological information in computer form means that we are likely to see an increase in informational patents in the future (Maschio and Kowalski, 2001). For example, in the field of systems biology vast quantities of disparate types of biological information are combined in computer models. Systems biology takes the computer-embodiment of biological information further than genomics to transcriptomes, proteomes, metabolomes and their interactions, as well as emergent properties and models. Patent applications and even some issued patents already cover these models (Allarakhia and Wensley, 2005). Arguably systems biology patents are better described as computer tools than biological information, but perhaps, as with the difficult distinction between the computer-embodiment of a genome and the genome itself, the distinction between computer tools and biological information is

becoming hard to establish in practice. As a consequence, the patent system may be increasingly stretched by these developments in contemporary biology.

Another scenario is that, if not a return to 'bricks and mortar' patenting, future patenting in the realm of biology might become solidly material again. Recent experimental studies have been conceptualised as producing 'synthetic genomes' (Pennisi, 2005). Rather than adding individual genes to an existing genome, these experiments aim to ‘redesign’ genomes in a more global sense. Even if it remains to be seen what can be achieved with synthetic genomes, such entities could be readily argued to be man-made and patented under the chemical regime of DNA patenting that was established with the *Chakrabarty*. In the future, the patenting of genomes may well run together with patenting in the field of nanotechnology, or as one report speculates, be "coupled" to insights of systems biology to make useful organisms (U.S. Department of Energy, 2004). If so, how would these entities fit into the U.S. Patent Classification?

With our study of three patent applications for whole genome sequences, we have shown how classification played a role in determining what kinds of entities were found to be patentable. In this way, the patent classification system will continue to play a role in the development of tomorrow’s technologies. It will be interesting to follow new types of biological patents – both informational and material – to see if they will force changes to the classification system as we saw with *Chakrabarty*, or whether the conservatism evident in our case study will prevail. It also remains to be seen whether complex informational models and synthetically modified nano-organisms will continue to be examined with respect to the categories provided by the existing classes of the U.S. Patent Classification, or whether new legal regimes, for example by drawing on other areas of the law, will be developed to establish ownership of such entities and inventions.

Acknowledgements

The authors would like to thank Maureen O'Malley, as well as Barry Barnes and John Dupré for commenting on versions of this paper, and Staffan Müller-Wille and Hyo Yoon Kang for their close reading and thoughtful comments. This research was part of the program of Egenis, the ESRC Centre for Genomics in Society at the University of Exeter. Adam Bostanci acknowledges a scholarship for doctoral studies from the University of Exeter.

References

Allarakhia, M. and Wensley, A. (2005). Innovation and intellectual property rights in systems biology. *Nature Biotechnology*, 23, 1485-1488.

Armistead, D. M., Fitzgibbon, M. J., Fleming, M. A., Griffiths, J. P., Kim, E. E., Kim, J. L., Sintchak, M. D., Thomson, J. A. and Wilson, K. P. (1999). Molecules Comprising a Calcineurin-Like Binding Pocket and Encoded Data Storage Medium Capable of Graphically Displaying Them. United States, Patent No. 5,978,740.

Beurton, P. J., R. Falk and Rheinberger, H.-J. (Eds.). (2000). *The Concept of the gene in development and evolution*. Cambridge: Cambridge University Press.

Blute, M. (2005). If the genome isn't a God-like ghost in the machine, then what is it? *Biology and Philosophy*, 20, 401-407.

Bult, C., White, O., Smith, H. O., Woese, C. R. and Venter, J. C. (1997). Complete Genome Sequence of the Methanogenic Archeon, *Methanococcus Janaschii*. Patent Application No. 08/916,412 (filed 22 August), US Patent and Trademark Office.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C. (1996). Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii. *Science*, 273, 1058-1073.

Cook-Deegan, R. (1994). *The gene wars: science, politics, and the human genome project*. New York, London: W. W. Norten & Co.

De Laet, M. K. and Mol, A. (2000). The Zimbabwe Bush Pump: Mechanics of a Fluid Technology. *Social Studies of Science*, 30, 225-263.

Eisenberg, R. S. (2000). Re-examining the role of patents in appropriating the value of DNA sequences. *Emory Law Journal*, 49, 783-799.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. and et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, 269, 496-512.

Fleischmann, R. D., Adams, M. D., White, O., Smith, H. O. and Venter, C. J. (1995). Computer Readable Genomic Sequence of *Haemophilus influenzae* Rd, Fragments Thereof, and Uses Thereof. Patent Application No. 08/476,102 (filed 6 June), US Patent and Trademark Office.

Fraser, C. M., Adams, M. D., Gocayne, J. D., Hutchinson III, C. A., Smith, H. O. and Venter, C. J. (1995). Nucleotide Sequence of the *Mycoplasma Genitalium* Genome, Fragments Thereof, and Uses Thereof. Patent Application No. 08/545,528 (filed 19 October), US Patent and Trademark Office.

Fraser, C. M., Gocayne, J. D., White, O. and et al. (1995). The Minimal Gene Complement of *Mycoplasma genitalium*. *Science*, 270, 397-403.

Godfrey-Smith, P. (2000). On the Theoretical Role of "Genetic Coding". *Philosophy of Science*, 67, 26-44.

Griffiths, P. E. (2001). Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science*, 68, 394-412.

Henry, J. W. (2004). Understanding and Simplifying Restriction Practice under 35 USC §121. *Journal of the Patent and Trademark Office Society*, 86, 17-73.

Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford: Stanford University Press.

Kevles, D. J. (2002). A history of patenting life in the United States with comparative attention to Europe and Canada. Luxembourg, Office for Official Publications of the European Communities.

Manis, J. J. (1981). Plasmid and process of isolating the same. United States, Patent No. 4,273,875.

Marshall, E. (1994). A Showdown Over Gene Fragments. *Science*, 266, 208-210.

Marshall, E. (1996). Patent Office Faces 90-Year Backlog. *Science*, 272, 643.

Maschio, T. and Kowalski, T. (2001). Bioinformatics - a patenting view. *TRENDS in Biotechnology*, 19, 334-339.

Moss, L. (2003). *What genes can't do*. Cambridge, Massachusetts: The MIT Press.

Myers, G. (1995). From discovery to invention: the writing and rewriting of two patents. *Social Studies of Science*, 25, 57-105.

Naini, A. A. (2006). Convergent Technologies and Divergent Patent Validity Doctrines: Obviousness and Disclosure Analysis in Software and Biotechnology. *Journal of the Patent and Trademark Office Society*, 86, 541-567.

Neumann-Held, E. M. (2006). The Genetic Code: Foundation of Life or Heuristic Device?. *Preprint of the workshop 'History and epistemology of molecular biology and beyond: problems and perspectives', held at the Max-Planck-Institute for the History of Science, Berlin, 13-15 October, 2005*, 141-149.

Nowak, R. (1995). Bacterial Genome Bagged. *Science*, 269, 468-470.

O'Malley, M. A., Bostanci, A. and Calvert, J. (2005). Whole-Genome Patenting. *Nature Reviews Genetics*, 6, 502-506.

Packer, K. and Webster, A. (1995). Inventing boundaries: the prior art of the social world. *Social Studies of Science*, 25, 101-117.

Packer, K. and Webster, A. (1996). Patenting Culture in Science: Reinventing the Scientific Wheel of Credibility. *Science, Technology, & Human Values*, 21, 427-453.

Pennisi, E. (2005). Synthetic Biology Remakes Small Genomes. *Science*, 310, 769-770.

Sarkar, S. (1996). Biological information: a skeptical look at some central dogmas of molecular biology. In S. Sarkar (Ed.) *The philosophy and history of molecular biology: new perspectives* (pp. 187-231). Dordrecht: Kluwer.

Shreeve, J. (2004). *The genome war*. New York: Knopf.

Sulston, J. E. and Ferry, G. (2002). *The common thread*: London: Bantam Press.

Toedt, D. C. (1995). Software as "Machine DNA": Arguments for Patenting Useful Computer Disks Per Se. *Journal of the Patent and Trademark Office Society*, 777, 275-289.

U.S. Department of Energy (2004). Synthetic Genomes: Technologies and Impact. http://www.sc.doe.gov/ober/berac/SynBio.pdf 6 January 2007.

U.S. Patent and Trademark Office (1996). Examination Guidelines for Computer-Related Inventions. *Federal Register*, 61, 7478-7492.

U.S. Patent and Trademark Office (2001). *Manual of Patent Examining Procedure*. Washington DC: United States Government Printing Office. Online at http://www.uspto.gov/web/offices/pac/mpep/ (last accessed 10/04/07).

U.S. Patent and Trademark Office (2005). *Handbook of classification*. Washington DC: United States Government Printing Office. Online at

http://www.uspto.gov/web/offices/opc/documents/handbook.pdf (last accessed 10/04/07).

Wade, N. (1995). Bacterium's Full Gene Makeup is decoded. *New York Times*. 26 May: 16.

---

[1] Philosophical efforts have gone into clarifying the gene concept (Beurton *et al.*, 2000; Moss, 2003). For the purpose of this paper, we understand the term 'gene' in a similar fashion to molecular biologists or patent practitioners who seek to identify protein-coding genes and pursue patent claims for Open Reading Frames, or ORFs. At the molecular level, genes are functional parts of the genome that are characterised by a particular DNA sequence. The argument developed in this paper does not require a tighter definition of what we mean by a gene.

[2] One might note, however, that other biological entities have been patented as computer-related inventions. For example, the 3D coordinates of the binding pocket of antibodies have been claimed as computer-systems for drug discovery (Armistead *et al.*, 1999).

[3] In addition to the genome patents discussed by O'Malley *et al.* (2005), perhaps an early precedent of genome patents are patents for plasmids as vectors for gene transfer (Manis, 1981).

[4] For the *H. influenzae* genome, we ordered the examination file for patent application number (08/476,102), and for *M. genitalium* and *M. jannaschii* we ordered the files for patent applications (08/545,528) and (08/916,412) respectively. Although these were not necessarily the original patent applications filed for each genome, it was possible to obtain the original applications, which were abandoned, to ascertain that the patent applications we studied were practically the same as the original applications. The patent examination files we chose had the advantage that they were associated with more correspondence between the patent attorneys and applicants. We ordered these files through a document supply service in Washington, DC.

[5] Patent applications are filed and prosecuted by law firms, and the process of patent examination can take several years. Due to staff turnover at the USPTO and at the law firm prosecuting the case, several different attorneys and examiners corresponded concerning each of the applications. For this reason, our references to 'the patent attorney' and to 'the patent examiners' can be read as idealisations of these actors as they emerge from our study of the patent examination process. If there are any face-to-face or telephone interactions of the patent examiners and applicants, then the nature of these interactions is recorded in the file, and in the files only contained notes only on a few trivial interactions. For this reason, we take the file as being a good reflection of the whole examination procedure.

[6] O. White was interviewed at The Institute for Genomic Research in Gaithersburg, Maryland, September 2004; R. Millman was interviewed by telephone at Alnylam Pharmaceuticals in March 2006. Interviews were carried out by A.B., the first author of this paper.

[7] This is now reclassified as *Methanocaldococcus jannaschii*.

[8] The patent arising from the *H. influenzae* genome patent application was issued on 4th March 2003 (US patent no. 6,528,289); likewise for *M. genitalium* on 25th March 2003 (US patent no. 6,537,773), and for *M. jannaschii* on 28th Sept 2004 (US patent no. 6,797,466).

[9] Possibly, the addition of the word 'fragment' was largely pro forma, as legal professionals often prefer established phraseology that has already been contested in previous case law.

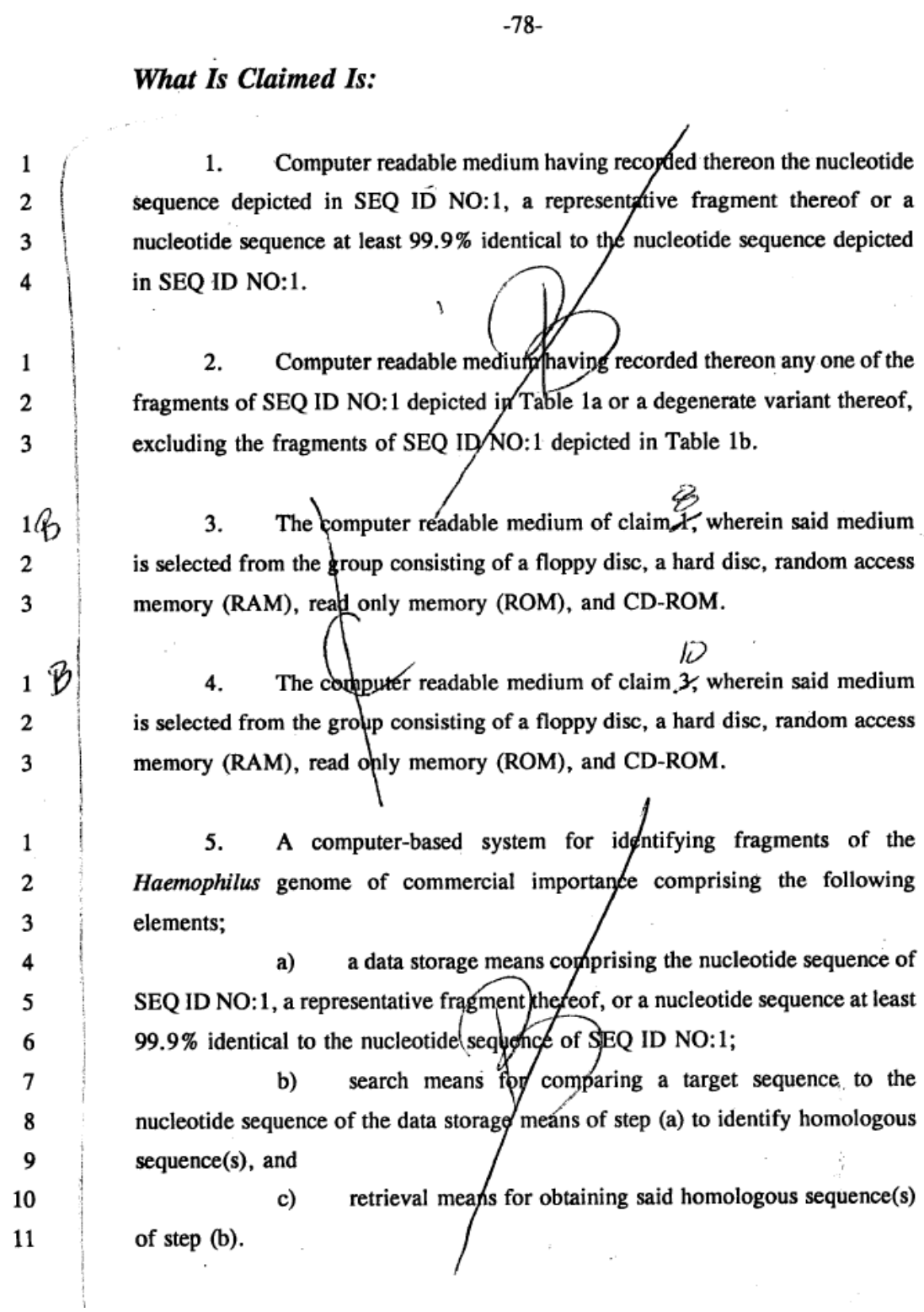
[10] Even if a claim for an isolated chromosome was formulated, one would also have to articulate a utility for this invention, which is not trivial, albeit vectors have been patented in this manner as tools for gene transfer.

[11] Our discussion has focused on the patenting criteria of utility and novelty. Taking DNA patenting as a starting point, however, one would not expect that the criterion of non-obviousness would present a hurdle to patenting a whole genome in the way envisaged by the patent attorney. Although it may be 'obvious' to determine the sequence of a DNA molecule of interest, whichever particular sequence one finds has long been held as non-obvious for the purposes of awarding biotechnology patents.

[12] That patenting practice presupposes distinct areas of prior art also manifests itself in the way the basic requirements for patentability are applied differently in different fields. The requirements of nonobviousness, enablement, and written description are applied in strikingly different ways with respect to biotechnology and software (Naini, 2006).

[13] The EPO, however, has the *ordre public* or public policy principle, which takes into account social and moral and economic values.

Figures:



-78-

What Is Claimed Is:

1. Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID NO:1, a representative fragment thereof or a nucleotide sequence at least 99.9% identical to the nucleotide sequence depicted in SEQ ID NO:1.

2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NO:1 depicted in Table 1a or a degenerate variant thereof, excluding the fragments of SEQ ID NO:1 depicted in Table 1b.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

5. A computer-based system for identifying fragments of the *Haemophilus* genome of commercial importance comprising the following elements;

a) a data storage means comprising the nucleotide sequence of SEQ ID NO:1, a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to the nucleotide sequence of SEQ ID NO:1;

b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and

c) retrieval means for obtaining said homologous sequence(s) of step (b).

Figure 1: The claims of the patent application for the genome of *Haemophilus influenzae* (USPTO Patent Appl. No. 08/476,102).

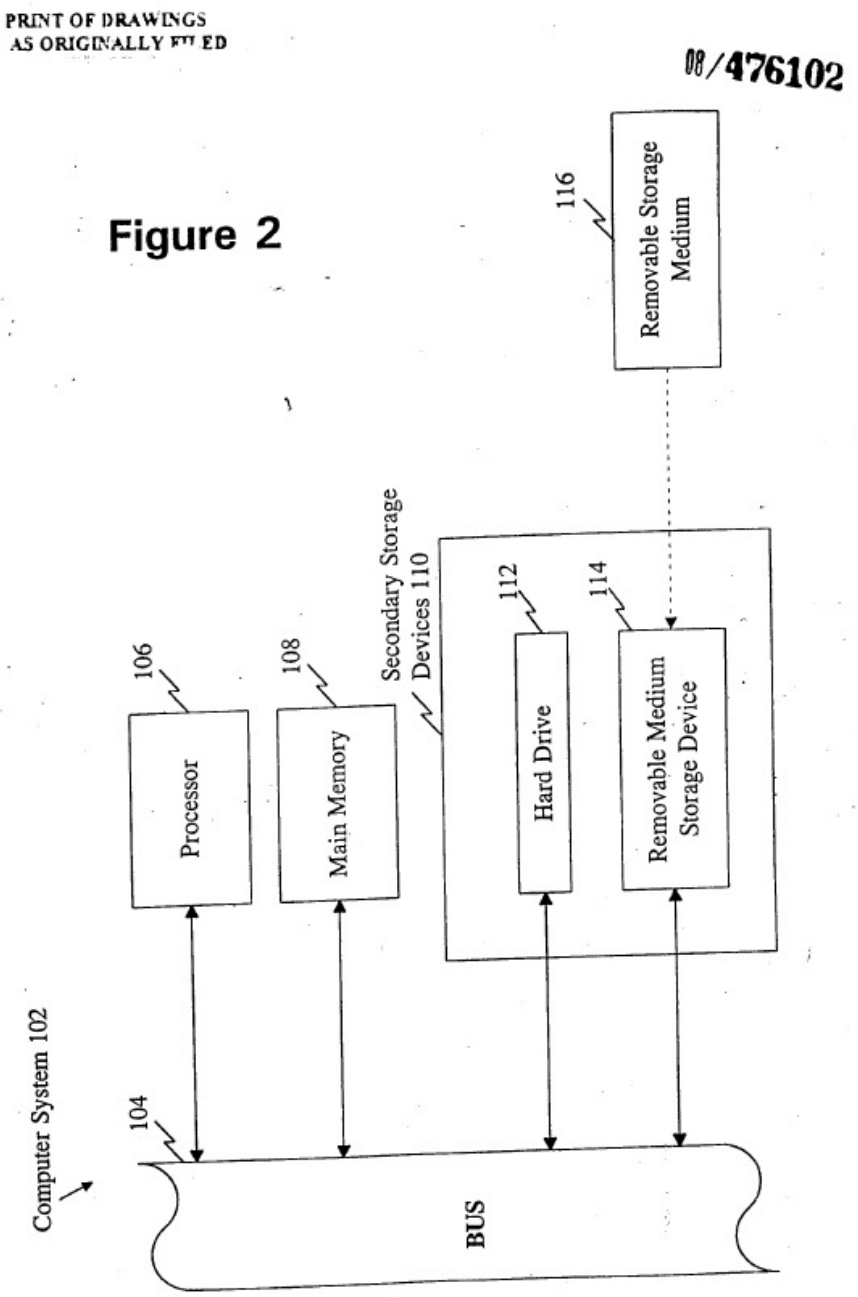


Figure 2: The genome as a 'Computer-Related' Invention. (USPTO Patent Appl. No. 08/476,102).